



LA-UR-24-28075

Author: Migeljan Imeri

Mentors: Brian Atkinson, Dominic Manno, Jason Lee

Title: Leveraging Lustre to Implement Incremental Indexing in GUFFI

Abstract: When dealing with filesystems on the order of terabytes or petabytes, trivial metadata operations such as looking up a file can suddenly become quite lengthy to perform. These lengthy metadata operations can negatively impact the user experience when they're attempting to manage their data and lower the performance of the file system as these operations are taking place. The Grand Unified File-Index (GUFFI) is a tool that indexes filesystems, allowing greatly improved performance when querying large parallel file systems. Currently to update this index, a full filesystem walk is necessary, which can be inefficient if only a small number of modifications are made in the filesystem. Using the Lustre changelogs feature, which records modifications to the filesystem, we can point GUFFI to reindex only specific directories, rather than having to do a full filesystem walk. Leveraging these changelogs should allow for significantly faster index updates. Faster index updates would allow for more up-to-date indexes, as we could perform these reindexes more frequently.